

# Weekly Report

2018.0827-2018.0902

## 1.This Week

### Power Grid Deep Learning Paper

- 1.We examined and processed the first batch of the 2,000 node dataset. Buses with missing values and samples without fault center information are filtered.
- 2.We generate a normal dataset to compare with the abnormal ones.
- 3.Despite from learning on which bus the fault takes place, experts are also interested in on which part of the bus the fault takes place (maybe on 2% location from bus A to bus B, or maybe on 98% location). As a result, we further examine the features of the samples that the fault happens on the same bus but on different parts of it (to see if it is possible to distinguish them). We project the raw samples to 2-dimensional space and find it undistinguishable. So we plan to process the data and compute features of it the next week.
- 4.We trained the model on the sampled dataset with missing values filled by 0. And we receive an accuracy of 69.8% on the test dataset, which proves that filling the missing data by 0 is not a good choice in our circumstance.
- 5.I read papers about automatic parameter tuning and parameter space visualization to find if it possible to do the suggestive visualization of parameter optimization.
- 6.Wu Fan went to Japan this week and will come back the next week.
7. Meetings summarization

会议	截止日期	备注
ICDE	每年两轮，一轮5月初，一轮10月初	支持的paper topic 有data visualization。但是这几年的论文中，几乎全部是计算、数据库的论文。（偏工程与计算）
ICDM	6月（2019的proposal还没出，以前都是6月due）	支持的paper topic基本都是data mining有关的。这几年论文中有部分跟机器学习、深度学习应用有关的论文
SIGKDD	2月	论文一般使用大数据，需要有挖掘到的insights。

### WaveLines Revision:

- 1.We carry out the evaluation process. The evaluation is divided into two parts:

- test the accuracy of users to find patterns (users are asked to find specific patterns in limited time)
- test the effectiveness of the system (users are asked to find possible patterns freely in limited time)

We gathered 15 results for all. The result of the accuracy test is very frustrating, most interviewee cannot reach an accuracy of 75% and we're trying to figure out this problem (if the test is too strict, we might need to redo the test again).

2. I revised the section of the evaluation process in the paper.

## **Others**

1. Organize the paper writing tricks (套路) gathered from others. Add the description of each paper type and the general construction of a paper. Extract the general sections (abstract, introduction, related work, discussion, conclusion) and organize them.

## **Working Hour: (except nap and eat time)**

**7 - 9 hours / week day**

**2 - 3 hours / weekend day (Thursday and Friday this week)**

**Total Working Hour this week: 45 hours.**

## **Paper Reading**

1. Long Short Memory Process: Modeling Growth Dynamics of Microscopic Social Connectivity (SIGKDD 2017)

This paper proposes a stochastic model (LSmP) to capture the growth dynamics of individual's social connectivity in a real world WeChat dataset (300 million users, 4.75 billion links). It firstly observes samples of individuals, and propose three hypothesis that preserves the feature of individual dynamics: average effect, multiscale effect, and correlation effect. These hypothesis are concluded into long-term (power law growth) and short-term memory (bursts) in the stochastic model. (And all have been proved by the author) All item in the model's formula is strictly proved and physically meaningful. In this paper, apart from the model itself, the findings (three hypothesis) are also strong contributions.

2. Visualization of Parameter Space for Image Analysis

3. VisTrails: Visualization meets data management

4. Comparative Visualization for Parameter Studies of Dataset Series

5. Preset based interaction with high dimensional parameter spaces.

These papers focus on optimizing parameter by applying visual approaches. Different combination of parameter settings are defined offline and computed offline to provide results for visually assessment. Offline computation could take a lot of time, so [2] proposes a parameter sampling method to reduce the computation cost. Few works have proposed online flexible parameter tuning plans and the problem is how to effectively sample suggestive parameter sets.

Visualization techniques uses in parameter space discovering includes simple linked charts, graphs [3], landscape [5], glyphs [4]. The key idea is to study the mappings of parameters -> outputs, and assess the outputs. But the assessment measures may vary greatly due to the change of the problem or the model. So a general approach is still missing.

## 2.Progress

Work	Deadline	Progress
<b>Power grid paper with Deeping learning</b>	-	1.Train the model on the 2,000 node dataset. 2.Test the model with different kind of data process methods.
<b>SQC Paper</b>	-	1.Talk to Zongzhuang the next week.
<b>WaveLine revision</b>	ASAP	1.Do the evaluation test and revise the paper.